DEPARTMENT OF HEALTH AND HUMAN SERVICES

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

For carrying out section 301 and title IV of the Public Health Service Act with respect to human genome research, [$382,384,000] *$426,739,000*.

[Departments of Labor, Health and Human Services, Education, and Related Agencies Appropriations Act, 2001. (P.L. 106-554)]

National Institutes of Health

National Human Genome Research Institute

Amounts Available for Obligation 1/

Source of Funding	FY 2000 Actual	FY 2001 Estimate	FY 2002 Estimate
Appropriation	$337,322,000	$382,384,000	$426,739,000
Enacted Rescission	(1,795,000)	(192,000)	---
Subtotal, Adjusted Appropriation	335,527,000	382,192,000	426,739,000
Real transfer to:			
Other NIH Institutes through the NIH Director's one-percent transfer authority	(70,000)	---	---
Other HHS Agencies through Secretary's one-percent transfer authority	(281,000)	---	---
HHS for the Office of Human Research Protection	---	(80,000)	---
Comparative transfer to:			
Other NIH Institutes as a result of a change in assessment formula for Central Services funding	335,000	---	---
Subtotal, adjusted budget authority	335,511,000	382,112,000	426,739,000
Unobligated balance lapsing	(47,000)	---	---
Total obligations	335,464,000	382,112,000	426,739,000

1/ Excludes the following amounts for reimbursable activities carried out by this account: FY 2000 - $9,121,000
FY 2001 - $42,900,000 FY 2002 - $45,710,000

Excludes $40,465 in FY 2000 and $104,642 in FY 2001 for royalties.

Justification

National Human Genome Research Institute

Authorizing Legislation: Sections 301, 485B, and 487(d) of the Public Health Service Act, as amended. Reauthorizating legislation will be submitted.

FY 2000 Actual		FY 2001 Estimate		FY 2002 Estimate		Increase or Decrease	
FTEs	BA	FTEs	BA	FTEs	BA	FTEs	BA
246	$335,511,000	260	$382,112,000	273	$426,739,000	13	$44,627,000

This document provides justification for the Fiscal Year 2002 activities of the National Human Genome Research Institute, including HIV/AIDS activities. A more detailed description of NIH-wide Fiscal Year 2002 HIV/AIDS activities can be found in the NIH section entitled "Office of AIDS Research (OAR).

INTRODUCTION

The National Human Genome Research Institute has the responsibility at the National Institutes of Health for providing leadership and support for the Human Genome Project and for conducting research aimed at understanding and treating both simple and complex genetic disorders. The Human Genome Project (HGP) is an international collaboration to characterize the complete set of genetic instructions encoded in the 3.1 billion base pairs of DNA. The HGP is funded in the United States by the NHGRI and the Department of Energy (DOE). The Project's international partners include the United Kingdom, France, Germany, Japan, and China.

We stand at a remarkable time in history. This year the HGP completed a working draft of the DNA sequence of the human genome. This clearly represents one of the most important milestones ever encountered on the never-ending quest to better understand ourselves and the wonder of life. To reach this milestone, the HGP's unique international collaboration drew upon the brilliant insights and painstaking work of more than a thousand scientists in laboratories all over the world. With remarkable speed and precision, these scientists managed to spell out, in the four-letter alphabet of DNA, nearly all three billion units of our genetic code. It is stunning to recall that not even 50 years have elapsed since the legendary team of James Watson and Francis Crick first discovered the elegant helical structure of DNA.

Tremendous challenges lie ahead, but with the profound new knowledge of our genetic sequence, humankind is on the verge of gaining immense and humbling new powers to heal. Genome science will revolutionize the diagnosis, prevention, and treatment of most human disease. In the coming decades, doctors will begin to cure diseases like Alzheimer's, Parkinson's, diabetes, and cancer by attacking their genetic roots.

While human genome sequencing is a flagship endeavor of the Human Genome Project, it is not the only goal. The 5-year NIH-DOE Human Genome Project Research Plan includes seven other ambitious goals that are guiding the development of a new and more diverse set of genomic research tools. These goals include: (1) optimizing current sequencing technologies and

development of novel strategies; (2) producing a catalog of common variations, or single-nucleotide polymorphisms (SNPs), in human DNA sequence; (3) developing new technologies and strategies for studying the function of genes and genomes; (4) completing the DNA mapping and sequence of additional model organisms; (5) crafting new approaches to addressing the ethical, legal, and social implications (ELSI) of research; (6) creating improved databases and analytical tools in bioinformatics and computational biology; and (7) establishing and nurturing training programs in scientific and ELSI aspects of genomic and genetic science.

The HGP is ushering in a new era of individualized preventive medicine that will improve human health. For its full potential to be realized, however, the accompanying ethical, legal, and social implications must be addressed. The NHGRI has been a leader in supporting rigorous scholarship and development of policy options for lawmakers to consider in their efforts to prevent the misuse of genetic information and protect the privacy of one's own genetic information.

The NHGRI also has a vibrant intramural research program. The Division of Intramural Research (DIR) has developed a cutting-edge scientific program to translate the tools of the Human Genome Project into knowledge about human genetic disease and its prevention, diagnosis and treatment.

The powerful research tools of the Human Genome Project and the research advances of NHGRI scientists are arming biomedical researchers with new information and techniques to unravel the mysteries of disease. This knowledge is dramatically accelerating the development of new strategies for diagnosis, prevention, and treatment of disease, not just for single gene disorders, but for the host of more common complex diseases (e.g., diabetes, heart disease, schizophrenia, and cancer) where genetic differences contribute to disease risk and response to particular therapies.

STORIES OF DISCOVERY

Sequencing the Human Genome: Our Genetic Instruction Book

June 26, 2000, marked the achievement of a historic scientific milestone - for the first time in history, humankind had available in public databases the initial reading of the text of our genetic instruction book, the three billion letters of DNA that constitute the human genome. It was a remarkable achievement and represented the tireless effort of thousands of dedicated scientists of the Human Genome Project working in twenty genome centers around the globe. The International Human Genome Sequencing Consortium published the sequence and initial analysis of the human genome on February 15, 2001 and revealed many remarkable insights. The Book of Life is actually at least three books. It's a history book - a narrative of the journey of our species through time. It's a shop manual, with an incredibly detailed blueprint for building every human cell. And it's a transformative textbook of medicine, with insights that will give health care providers immense new power to treat, prevent, and cure disease. This achievement represents just the end of the beginning for the

The Human Genome Project has, from its beginning, enjoyed remarkable success. Many of the project's initial goals have been achieved, including building maps to localize genes in both the human and mouse genomes, and sequencing the genomes of model organisms including the

bacterium *E. coli*, baker's yeast, fruitfly and the roundworm *C. elegans*. Building on those successes, this year HGP scientists reached a historic milestone: the completion of a working draft sequence of the human genome.

On June 26, 2000, leaders of the public Human Genome Project and Celera Genomics Corporation jointly announced that both groups had completed separate working drafts of the DNA sequence of the human genome. This historic scientific milestone was also celebrated by the hundreds of scientists throughout the world who contributed to the public Human Genome Project as members of the International Human Genome Sequencing Consortium.

The Human Genome Project and Celera Genomics used different, but complementary, approaches to sequence the human genome. The HGP's strategy for sequencing the human genome involved two tasks: placing large fragments of cloned DNA in the proper order to cover all of the human chromosomes, and determining the DNA sequence of these fragments. Celera used a "whole genome shotgun" approach, in which the entire genome is shredded into small fragments that are sequenced and put back together on the basis of sequence overlaps. Celera utilized the public data extensively; more than half of their sequence data came from public databases. They also discovered that the pure whole genome shotgun approach may not be suitable for the sequencing of a mammalian genome because of the very high repeat content.

Between March 1999 and June 2000, the production of human genome sequence skyrocketed. During this time, Human Genome Project scientists sequenced 1,000 DNA letters a second--7 days a week, 24 hours a day. The resulting working draft sequence covers over 94% of the human genome with over one-third in highly accurate finished form. The average accuracy of all of the DNA sequence in this assembly is 99.9 percent.

In the months following the completion of the working draft, Human Genome Project scientists - including many of the most brilliant genetics and computational experts in the world - began to scour the sequence to find answers to age-old questions in biology. Indeed they found surprising answers to age old mysteries in biology; they also unrevealed whole new mysteries that will occupy the minds of researchers for years to come.

Among the treasure trove of discoveries from this initial analysis were:

1) Humans have only 30,000-35,000 genes, far fewer genes than the 80,000-150,000 that had been predicted.

2) Individual human genes are able to produce multiple different proteins.

3) The structure of human genes is much more complex than those of simpler organisms.

4) Over 200 genes were found in the human genome that appear to have come from bacteria.

5) The mutation rate in males is twice that found in females.

6) Much of the "junk" in the genome appears to have important functions.

Because of the enormous value of sequence information to researchers around the world, HGP scientists have placed all DNA sequence data in public databases where it is immediately and freely available with no restrictions on its use or redistribution. The information is scanned daily

by tens of thousands of scientists in academia and industry, as well as by commercial database companies providing information services to biotechnologists. More information about the sequencing and analysis of the human genome is available at: http://www.nhgri.nih.gov/genome_sequence.html

Genome-scale Analyses: cDNA and Tissue Chips

The goals of the Human Genome Project include far more than sequencing the genomes of humans and model organisms such as the laboratory mouse. The HGP is revolutionizing the way biology and medicine will be explored in the 21st century and beyond. The HGP is focused on developing tools and training researchers to exploit the voluminous set of genetic information now available. The availability of entire genome sequences is enabling a new approach to biology often called functional genomics, the interpretation of the function of DNA sequences on a genomic, or holistic, scale. This involves the comprehensive analysis of gene expression.

The newfound abundance of genomic information is propelling scientists out of the pattern of studying genes individually; now, scientists are able to monitor thousands of genes at a time. For such large-scale analyses, miniaturized "DNA chip" technologies, also called microarrays, can be rapid, efficient, and economical. All microarrays share this characteristic: they permit researchers to examine many elements in parallel. NHGRI has supported the development and application of the major microarray technologies in use, including chips sold commercially, and has promoted communication among researchers in the field.

The flood of data emerging about DNA and genes has required new ways of sorting through the information to find the telling details that will illuminate how the cells that make up living organisms function. Microarrays are being used for many different applications. Some microarrays gauge how active different genes are in different kinds of cells; others let researchers track the molecular changes in tumor cells as cancer progresses. Capturing holistic views of changes within cells has begun to elucidate the signaling pathways that are altered and distinguish a cancerous cell from a non-cancerous cell. Such insights may provide the identification of early, presymptomatic changes in cells and thus, rational therapeutic targets for the treatment of cancer.

Expression arrays, which chart gene activity, have been among the most productive chips so far. To make an expression array, robots spot fragments from thousands of genes onto a single glass microscope slide about the size of a postage stamp. This genomic approach to understanding the impact of an altered genetic code within a cell is analogous to listening to a full orchestra playing a symphony, rather than individual instruments one at a time. While each cell contains the score for each of the approximately 35,000 or so available instruments (the protein products of genes), it only utilizes a fraction of those. Each different cell type (e.g., brain or muscle cell) will utilize a different complement of 10,000 or so instruments. While much can be learned from studying an individual instrument, significant limitations exist. Our knowledge grows dramatically when we begin to examine the networks of pathways that are affected by a single alteration in the genome.

In the first large-scale cancer genetics studies to be made possible by the wealth of information generated by the international Human Genome Project, NHGRI researchers have used expression microarrays to discover genetic "signatures" that can differentiate between different types of melanoma and to easily distinguish between hereditary and sporadic forms of breast cancer. Such classification of cancer on a molecular level offers the possibility of more accurate diagnosis and prognosis of a particular patient's tumor, based on his or her genetic makeup. It also offers the hope of tailoring therapies to the individual.

In order to determine the importance of any gene in a more physiological setting, a second kind of array, called the tissue microarray, can confirm the importance of each gene that emerges as a candidate. NHGRI researchers have developed a way of arranging some 1000 tiny cylindrical tissue biopsies in a small paraffin block. Thin slices cut from this block can be mixed with a probe that binds to a specific gene or gene product to allow researchers to visualize gene number, activity or subcellular localization of proteins in hundreds of different tissues simultaneously. Tissue arrays permit researchers to examine the molecular details of many different healthy tissue types or in different stages of disease. This past year, NHGRI researchers combined cDNA and tissue microarray technologies to identify molecular alterations associated with the progression of human breast cancer.

First, researchers applied their new double-chip approach to human breast cancer cells grown in culture to identify genes that were over expressed. Then, tissue chips were constructed, containing tiny dot-sized samples from over 600 breast cancer tumors, and used to determine how frequently these genes were over-expressed in cancers from actual patients. One of the amplified genes, HER-2, on human chromosome 17, had previously been shown to be involved in the initiation and progression of breast cancer. As expected, patients whose tumor samples had amplified HER-2 had a poor survival. However, the researchers were surprised to discover on chromosome 17 another gene that was also over-expressed. This gene, S6K, is know to be involved in cell proliferation and thus represents a good candidate for a cancer-related gene.

In all their various forms, microarray technologies can support the study of genetic complexity and are becoming increasingly common as a “genome perspective,” one that considers the entire DNA code of an organism, takes root in biomedicine.

Single Nucleotide Polymorphisms (SNPs): New Tools for Tracing Inherited Diseases

The DNA sequence of any two people is 99.9% identical. The 0.1% difference includes genetic variation that has no effect at all, others that make each of us unique, and some that are associated with differences in the risk of getting various diseases or having an adverse drug reaction. Some diseases, such as cystic fibrosis and Huntington’s disease, result from differences in DNA sequence in single genes. However, many common diseases such as diabetes, cancer, heart disease, psychiatric disorders, and asthma are influenced by complex interactions between multiple genes as well as by non-genetic factors such as diet, exercise, smoking, and exposure to toxins. With the tools of the Human Genome Project, identifying the genes for diseases caused by alterations in single genes has become relatively straightforward. Finding the genes that contribute to common diseases remains

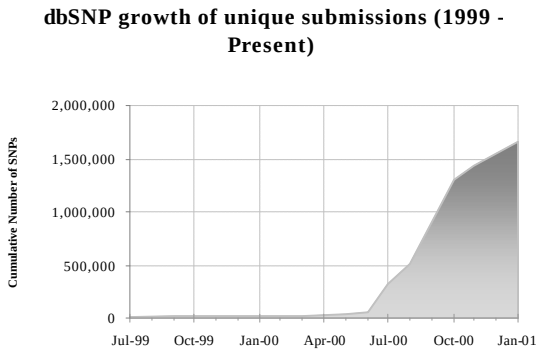
A key aspect of research in genetics is associating sequence variations with heritable diseases. The most common variations are single nucleotide polymorphisms (SNPs). These are "spelling" variations in the DNA code; between any two individuals there will be about one such variation every 1000 bases. Because SNPs are expected to facilitate large-scale genetic association studies, there has recently been great interest in SNP discovery and detection. The identification of SNPs has accelerated dramatically in the past year, due in large part to the availability of working draft sequence of the human genome.

In FY 1999, NHGRI organized the establishment of the DNA Polymorphism Discovery Resource (PDR) to allow researchers to look for SNPs in a standard set of samples. The PDR is now the major resource being used to look for SNPs. It consists of 450 DNA samples collected under strict ethical guidelines from anonymous unrelated United States residents of diverse ethnic backgrounds (http://www.nhgri.nih.gov/About_NHGRI/Der/variat.htm)The NHGRI has funded studies to allow researchers to look for SNPs in this set of samples. This permits the accumulation of a data set that can then be used in association studies to identify inherited disease risks.

This effort is complemented by The SNP Consortium (TSC), a non-profit entity whose mission is to develop a high-quality SNP map of the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions (http://snp.cshl.org/). The project started in April 1999 and is anticipated to continue until the end of 2001. The SNP Consortium's members include the medical research charity The Wellcome Trust; 10 pharmaceutical companies including AstraZeneca PLC, Aventis Pharma, Bayer AG, Bristol-Myers Squibb Company, F. Hoffman-La Roche, Glaxo Wellcome PLC, Novartis, Pfizer Inc, Searle (now part of Pharmacia), and SmithKline Beecham PLC; Motorola, Inc.; IBM, and Amersham Pharmacia Biotech. Academic centers including the Whitehead Institute for Biomedical Research, Washington University School of Medicine in St. Louis, the Wellcome Trust's Sanger Centre, Stanford Human Genome Center, and Cold Spring Harbor Laboratory, are involved in SNP identification and analysis.

In July 2000, the HGP and TSC announced a collaboration to accelerate the construction of a higher-density SNP map and enhance the utility of human working draft sequence. At the same time, the data generated will help improve the "working draft" itself. Three academic genome research centers - the Whitehead Institute for Biomedical Research, Washington University School of Medicine, and the Sanger Centre - participated in this collaboration.

Through this collaboration, The SNP Consortium has been able to contribute about three times as many SNPs to the public domain as otherwise would have been possible under TSC's original scientific plan which had been to identify 300,000 SNPs and map at least 150,000 of these SNPs, evenly distributed throughout the genome. An exponential increase in the amount of human genetic sequence data that became available from the Human Genome Project during FY 2000 has enabled the consortium to proceed at a much faster pace than originally envisioned.



As of February 12, 2001, the public database that serves as a central repository for SNPs, dbSNP (http://www.ncbi.nlm.nih.gov/SNP/index.html), had received submissions for a non-redundant set of 1,653,181 SNPs for the human genome.

SCIENCE ADVANCES

Researchers Decipher the First Two Chapters of the Human Genetic Instruction Book

The human genome is packaged into 23 pairs of chromosomes, often referred to as chapters in this vast genetic instruction book. While individual genes have been identified and sequenced for decades, up until the past year, scientists never had the opportunity to look at the genomic landscape of an entire chromosome. It has been likened to seeing an ocean liner emerge out of the fog, when all you've ever seen before were rowboats.

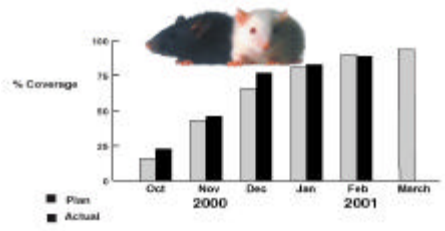
In the December 2, 1999, issue of Nature, an international team of researchers reported for the first time the complete sequencing of chromosome 22. Just a few months later, in the May 18, 2000 issue of Nature, scientists in Japan and Germany published the essentially complete genetic sequence of human chromosome 21. Genes on chromosome 21 are involved in Down syndrome as well as other disorders, including Alzheimer disease, certain cancers, and manic depressive illness. Chromosome 22 is implicated in the workings of the immune system, congenital heart disease, schizophrenia, mental retardation, birth defects, and several cancers including leukemia.

Sequencing the Mouse Genome: Providing tools to interpret the Human Genome and a powerful model system

Deciphering the genomic sequence of model organisms is a critical component of the interpretation of the human genome. A unique strength of human-mouse sequence comparisons is their ability to reveal the parts of the genome that code for protein, and also some of the regulatory sequences that are involved in turning genes on and off. The mouse is one of the most significant laboratory animals for studying human disease because mice and humans share many of the same fundamental biological and behavioral processes, and a great number of mouse models of human diseases have already been identified. Mouse models provide scientists with unprecedented insights into the molecular basis of disease and the response to potential therapeutic agents.

The effort to sequence the genome of the laboratory mouse was launched by NHGRI in October 1999. NHGRI's initial efforts focused on achieving broad coverage of the genome while completing maps that establish the physical organization of the 21 mouse chromosomes.

The Mouse Sequencing Consortium (MSC) - comprising three private companies, six Institutes of the NIH and the Wellcome Trust - was formed in October 2000 to work collaboratively to produce a publicly accessible draft sequence of the mouse genome in six months. This public-private effort to accelerate the sequencing of the mouse genome has exceeded its own goals and expects to reach its target of three-fold coverage by April 2001. In the second stage, attention will turn to filling any gaps in the draft and finishing the sequence in high quality, final form no later than 2005.



Rat genome sequence will provide significant gains in human pharmacological studies

In February 2001, NHGRI and the National Heart, Lung and Blood Institute announced an expansion of the NIH program to determine the DNA sequence of the genome of the laboratory rat, a key experimental animal for many areas of biomedical research, including cancer, high blood pressure and pharmacology. The work will be carried out be a joint effort between Baylor College of Medicine, Celera Genomics and Genome Therapeutics Corporation. The dual strategy being employed for sequencing the rat genome will take advantage of the lessons learned during the initial sequencing of the human genome. Data from the project will be released weekly into public databases. This step represents the latest in several positive efforts to integrate the talents of the public and private sectors in genomic sequencing.

Mammalian Gene Collection: A Resource for Studying Gene Expression and Function

When a gene is active in a cell, the functional, or coding part of the gene is "transcribed" into a mRNA molecule, which is in turn, directs synthesis of a protein. A particularly powerful material for studying gene expression and gene function, therefore, is cDNA, which represents the full-length, expressed mRNA transcript. Indeed, one of the most effective and widespread manifestation of the genomics revolution has been the ready public access to cDNA libraries, sequences, and clones.

In FY 1999, the NHGRI and the National Cancer Institute, with participation of 15 additional NIH Institutes, launched an initiative on full-length mammalian cDNAs, the Mammalian Gene Collection. The project's scientific goals are to produce publicly accessible cDNA clone collections that contain full-length copies of all genes, to sequence these cDNAs, to develop the associated informatics tools, and to create a publicly accessible website to provide up-to-date information to the research community. Already nearly 20,000 full length cDNAs have been identified and prioritized for full length sequencing.

Fruit fly genome sequence will provide further insights to cancer and aging

Drosophila melanogaster, more commonly known as the fruit fly, has been a powerful model system in biology since the early the 20th century. *Drosophila* studies have played a pivotal role in research ranging from aging and cancer to learning and memory. The sequence of the euchromatic portion of the genome of *Drosophila* is close to completion. Scientists at Celera Genomics, the University of California at Berkeley, and the Baylor College of Medicine carried out the bulk sequencing. The sequence data are available through GenBank, and the annotation is available through the GenBank and FlyBase databases. The combination of whole genome shotgun sequencing with a mapped BAC-by-BAC approach was effective, yielding ~120 Mb of data containing ~2400 gaps. NHGRI continues to fund the Berkeley and Baylor efforts to close the remaining gaps, to elucidate complex repeat regions, and to ensure the finished sequence meets quality standards for finished data. Approximately 400 gaps remain at this time.

Center for Inherited Disease Research: A service to help researchers identify genes that contribute to human disease

The Center for Inherited Disease Research (CIDR), located on the Bayview campus of The Johns Hopkins University, provides high-throughput genotyping services, study design advice, sophisticated data warehousing technologies and database assistance to research efforts

attempting to identify genetic loci and allelic variants involved in human disease. CIDR is fully funded through a $21 million federal contract from the National Institutes of Health to The Johns Hopkins University. CIDR is a joint effort by twelve NIH institutes with the National Human Genome Research Institute serving as the lead agency.

To date, 66 projects (out of a total of 165 different projects requesting access) have been accepted for genotyping at CIDR, including studies of colon cancer, lung cancer, schizophrenia, Alzheimer's disease, NIDDM, bipolar disorder, non-syndromic deafness, obesity, hereditary nonpolyposis colorectal cancer (HNPCC), osteoporosis, and dystonia. The current genotyping capacity of CIDR is over 5.0 million per year and is projected to increase to 8.5 million genotypes per year by the end of FY 2002.

The services provided by CIDR are expediting the identification of genes involved in a variety of diseases. This is a crucial step in understanding the molecular basis of disease and is an early, but vital step in the development of improved diagnostic and treatment strategies.

Bringing the Human Genome Project into the Classroom
A new multimedia educational tool for students and the public.

Knowledge of the human genome is increasing exponentially. The application of this knowledge to improve health and medicine and expand our understanding of ourselves as a species and as communities, families, and individuals will be profound. To maximize the benefits and to ensure informed public and personal decision making about science, medicine and policy, the public must understand basic concepts of genomics and genetics. To ensure that science teachers throughout the country have better access to the latest information about the Human Genome Project, NHGRI has created a free multimedia kit, entitled "The Human Genome: Exploring our Molecular Selves".

The kit is designed to educate, engage, and excite users about the human genome and genetics. The primary target audience is high school students, though a much broader use with college students, voluntary health organizations and the general public is anticipated. This project is being co-sponsored by the NIH, Department of Energy, Howard Hughes Medical Institute, Pharmaceutical Research and Manufacturers of America (PhRMA), and the journals Science and Nature. The kit was released concurrently with the publication of the human sequence analysis in February 2001.

The kit includes a multimedia CD-ROM, wall poster of the genome, an informational brochure and a video documentary that was awarded a gold medal by the New York Film Festival.

Scientists Pinpoint Location of Possible Third Gene
Involved in Hereditary Breast Cancer to Chromosome 13

Researchers in Finland, Iceland, and Sweden, working with scientists at the NHGRI have found evidence of a new gene that appears to increase susceptibility to hereditary breast cancer. The study examined women who live in Nordic countries and who have three or more female family members with breast cancer. The finding may help explain why some women with a family history of hereditary breast cancer are at particularly high risk of developing the potentially fatal disease, even when they lack mutations in two previously identified breast cancer susceptibility genes, BRCA1 and BRCA2.

Genetic Mutation Causes Common Defect in Early Development of Human Forebrain

An international team led by scientists at the NHGRI located one of the genes that can cause holoprosencephaly, the most common structural defect of the developing forebrain in humans. It results in varying degrees of mental retardation. The finding suggests that the gene, TG-interacting factor (TGIF) plays an important role in the brain's separating into left and right hemispheres during fetal development. The TGIF gene is the fourth found in humans to be involved in HPE.

Gene Linked to Developmental Syndrome in Old Order Amish

Scientists at NHGRI identified an altered gene responsible for a rare, recessive developmental syndrome found predominantly among the Old Order Amish population. Called McKusick-Kaufman syndrome, or MKS, the condition is the first human disorder to be attributed to a mutation in a gene affecting a type of molecule called a chaperonin.

OTHER PRIORITIES

Ethical, Legal, and Social Implications of Human Genome Research (ELSI)

The NHGRI recognized from its inception its responsibility to address the broader implications of newfound abilities to decipher genetic information. The NHGRI commits 5 percent of its annual research budget to study the “ethical, legal, and social implications (ELSI)” of genome research. The ELSI Research Programs of the NHGRI and the Department of Energy sponsored a conference in January 2001 to reflect on the past, present, and future of ELSI research and to consider its impact on genetic research, health and public policy. The ELSI program has now completed its eleventh year, having spent just over 70 million dollars, funded about 200 research and education projects, and generated about 450 grant products in the form of peer review publications, videos, curricula, etc.

A panel of outside experts, the ELSI Research Planning and Evaluation Group (ERPEG), recently released its report, “A Review and Analysis of the ELSI Research Programs at NIH and DOE”. ERPEG concludes in this report that the ELSI research programs at NHGRI and DOE have succeeded in establishing a new and vital field of research that has laid a strong foundation for future research in this area. They also made a series of recommendations for enhancing the ELSI programs.

The study of human genetic sequence variation and the exploration of related ethical, legal and social issues were identified in 1998 as two of the new five-year goals of the Human Genome Project. To that end, in April 1999, the ELSI program released an RFA entitled “Studies of the Ethical, Legal and Social Implications of Research into Human Genetic Variation.” Of the twelve applications received in response to the RFA, five were funded in FY 2000. These projects address a range of topics, including: (1) a comparative analysis of the processes for obtaining informed consent to genetic variation research in Africa and the United States; (2) stigmatization in the context of genetic research; (3) pharmacogenomics and minority populations; (4) risks and benefits of genetic research in Native American communities; and (5) the impact of genetic variation research on conceptions of ethnicity, citizenship, and family.

Genetic Screening for Hereditary Hemochromatosis: In FY 1999, NHGRI and NHLBI co-sponsored an RFP to fund contracts to carry out an epidemiological study of the prevalence, genetic and environmental determinants, and potential clinical, personal and societal impact of iron overload and hereditary hemochromatosis (HH) in a multi-center, multi-ethnic, primary care based sample of approximately 100,000 adults. This is a five-year, $30 million project.

Contracts to carry out this project were awarded in early FY 2000 to five field centers, a coordinating center, and a central lab. A Steering Committee was formed and met three times in FY 2000 to come to consensus on the overall study protocol; subcommittees were also established and worked throughout the year to develop protocols for each stage of the study. Recruitment is beginning in FY 2001.

The portion of the study NHGRI is funding will examine the ethical, legal, and social issues related to the possibility of implementation of primary care-based screening for iron overload and HH, including identification of appropriate health care delivery models and the potential personal, societal, or family-related impact of and barriers to primary care- or population-based screening and genetic testing. This information will be valuable in the future development of recommendations regarding genotypic and/or phenotypic screening not only for HH, but also for other potentially treatable adult-onset genetic disorders.

Privacy and Fair Use of Genetic Information: In a Time/CNN poll conducted in June 2000, 75% of those polled indicated they would not want their health insurance company to have information about their genetic code. Another Time/CNN poll in 1998 revealed that 95% of those polled thought employers should not be able to access the genetic record of their employees without permission. This is evidence of public concern that affects the choices individuals make about their own health as well as their decisions about whether to participate in research. In genetic testing studies at the NIH, nearly 32 percent of eligible people offered a test for breast cancer risk decline to take it. The overwhelming majority of those who refuse cite concerns about health insurance discrimination and loss of privacy as the reason.

Workshops conducted by the NIH-DOE ELSI Working Group and the National Action Plan on Breast Cancer (NAPBC) formulated recommendations to protect individuals from genetic discrimination in health insurance, genetic discrimination in the workplace and to enhance the privacy of genetics research data. Many of these recommendations have been used by federal and state legislators in crafting legislation.

Current protections, both at the federal and state levels include:

- The Health Insurance Portability and Accountability Act of 1996 (HIPAA) provides important protection from genetic discrimination by group health insurers. Genetic information may not be used to deny coverage and cannot be considered as a preexisting condition, in the absence of a diagnosis of the condition.
- On February 8, 2000, an Executive Order was signed that prohibits genetic discrimination in employment by agencies or departments of the Executive branch. The Executive Order also contains provisions to protect the privacy of genetic information.

- Thirty-eight states have enacted legislation that addresses genetic discrimination by health insurers. Some of this legislation also contains provisions to protect the privacy of genetic information.
- Twenty-six states have enacted legislation that addresses genetic discrimination in the workplace. Some of this legislation also contains provisions to protect the privacy of genetic information.

Despite these important federal and state activities, gaps in protection from genetic discrimination persist. The ELSI program at the NHGRI will continue to work on genetic discrimination and privacy issues and to forge consensus recommendations that balance the need to protect individuals, the needs of the research community, and the needs of the healthcare industry to utilize genetic information.

Health Disparities in African-Americans

During the past several years the Office of Research on Minority Health (now the National Center for Minority Health and Health Disparities) has supported an innovative research collaboration between investigators from Howard University and scientists in the intramural research program of the NHGRI. The collaboration involves support for projects involving African-American individuals affected with breast cancer, diabetes, or hereditary prostate cancer.

The goal of these studies is to establish a Center at Howard University for collaborative research on genomic analyses of diseases that disproportionately affect African Americans. In addition, Howard University and the NHGRI are serving jointly as research-training sites for African-American scientists involved in these projects.

*Africa America Diabetes Mellitus Study (AADM)**:*** Because of the high frequency of environmental risk factors for diabetes in the African-American population, it is potentially more powerful to study genetic risk factors in West Africans, since they are thought by many anthropologists to be the founding population of modern African-Americans and have fewer dietary and nutritional confounding variables. Five recruitment sites in Nigeria and Ghana were selected through a peer review process. Patient samples and other clinical data are sent to the Coordinating Center at Howard University. Based on the successful recruitment of study participants during a one-year pilot project, a full-scale study was implemented in September 1998 with an anticipated total of 400 pairs of siblings affected with diabetes mellitus. This goal was met in the fall of 2000. Genotyping of samples from West Africa is currently underway at the Center for Inherited Disease Research in Baltimore. The study has not only started to yield high quality data, but has assisted in the recruitment of several top-flight scientists to the Center at Howard University.

African-American Hereditary Prostate Cancer Study Network (AAHPC): The Howard Center is also coordinating a linkage study of hereditary prostate cancer. The initial aim is to enroll 100 families with prostate cancer in which at least four men, diagnosed at or before 65 years of age, are affected in each family and there are four other (unaffected) relatives available for study. African-American prostate cancer families of this description are almost completely missing from other pedigree collections, despite the higher incidence and higher lethality of prostate cancer in black men. Already more than 40 families have been identified and genotyping has begun. Blood samples and clinical data are sent from recruitment sites around the country to the

Howard Center for DNA extraction. Samples from these families are being studied to see if linkage can be found to a known hereditary prostate cancer location on Chromosome 1 as well as to determine if other linkages exist.

Second Annual NHGRI Consumer Day

NHGRI hosted the Second Annual NHGRI Consumer Day on November 9, 2000. The goal of this event was to increase awareness of NHGRI, the Human Genome Project and genetic/genomic research. Another vital goal of this effort was to establish and strengthen ongoing relationships between NHGRI and those attending Consumer Day, particularly the members of voluntary health organizations. Scientists, consumers and scholars led presentations on topics such as genetic variation, fair use and privacy of genetic information, participating in research and genetic testing and counseling. NHGRI worked with leaders of consumer and professional organizations to plan and promote this event. Over 200 participants attended Consumer Day 2000. Feedback was very enthusiastic and NHGRI is planning its 3rd Annual Consumer Day program for fall 2001.

Genomics Short Course for Faculty at Minority Institutions

The NHGRI intramural program hosted its 2000 *Genomics Short Course for Faculty at Minority Institutions* (August 8-11, 2000). The annual Short Course updates college faculty from institutions with substantial minority enrollment on the latest developments in genetic technology, medical genetics, gene therapy and ethics. The course also assists attendees in incorporating this information into classroom teaching to cultivate minority student interest in genome research, and offers information on careers in genetics and grant writing skills. Participants visit NHGRI laboratories and experience firsthand the latest technologies and research. A total of 27 faculty members from national universities and colleges with substantial minority enrollment participated in the program this year.

HIVAIDS-Related Research

In its intramural program, NHGRI is pursuing strategies to prevent or cure AIDS. Ultimately the successful use of intracellular immunization to combat HIV infection will require the integration of virus vectors containing efficient anti-HIV elements into the adult Hematopoietic Stem Cell (HSC), the ultimate progenitor of all peripheral blood cells. If an anti-HIV element can be introduced into adult HSC, it will be passed along to all of the progeny of that adult HSC, ensuring the continuous, life-long production of HIV resistant cells that would permanently protect the patient from HIV spread.

The successful modification of adult HSC by viral vectors has three important steps. First the viral vector must bind to a specific receptor on the surface of the adult HSC. Researchers in the NHGRI intramural program have developed methods to purify human adult HSC and have determined that the receptors for the retroviral gene transfer vectors currently in use are not present on adult HSC but receptors for several viruses are abundant. These viruses are now being analyzed for use to introduce anti-HIV elements. The second important step in the modification of adult HSC requires the adult HSC to divide so that the new anti-HIV sequences can become integrated into the DNA of the target cell. NHGRI intramural researchers plan to use highly purified human adult HSC to determine the optimum combination of growth factors that will promote adult HSC division during the exposure to virus vectors, while preserving the ability of the adult HSC to generate a life-time supply of red and white blood cells.

The final important step for successful modification of adult HSC with anti-HIV elements is that the anti-HIV elements have to be produced in the mature progeny of the adult HSC at all times so that the cells are always prepared to interrupt HIV infection. Many groups have shown that viral vectors can become "silenced" over time and stop making the critical elements. To combat silencing, NHGRI intramural investigators are developing and evaluating different virus vectors that contain genetic elements to prevent silencing.

NHGRI will also initiate a genetic approach to the study of AIDS that takes full advantage of the strengths of the intramural NHGRI clinical research program. These approaches are aimed at understanding the genetic component of innate HIV resistance. NHGRI clinical researchers will recruit and evaluate families in which HIV positive individuals without symptoms of AIDS have been identified. As larger numbers of families are recruited, the NHGRI Inherited Disease Research Branch will use powerful gene mapping and statistical genetics analyses to identify additional genetic loci that contribute to HIV resistance. It is anticipated that the results of these activities will provide novel ideas to evaluate for the treatment and prevention of AIDS.

NEW ACTIVITIES

Centers of Excellence in Genomic Science

In FY 2001, NHGRI awarded the first grants under its new program, Centers of Excellence in Genomic Science (CEGS). This program stimulates the development of new genomics approaches and their integration into biomedical research. It also expands training opportunities for a new generation of research scientists and bioinformaticians. New groups of investigators who have not collaborated previously are coming together, breaking down traditional academic boundaries across biology, chemistry, physics, computer science, math, statistics, and medicine, to devise new conceptual approaches for comprehensive genomic data collection and analysis.

Additional centers will be funded in FY2002 to focus on new approaches to genome-wide analysis in areas such as: regulation of gene expression; protein expression and interaction; gene and protein networks; human genetic variation related to health and disease. Other centers will focus on the storage, analysis, and distribution of comprehensive data sets; and the collection of high quality data to support these scientific analyses in the most cost-effective manner possible. These examples notwithstanding, the real strength of the CEGS program is that investigators may submit completely new ideas that we cannot even anticipate, in this exciting emerging area.

Model Organisms and Comparative Sequencing

With the working draft of the human genome completed, the NHGRI large-scale genomic sequencing program moves on toward capitalizing on the extensive sequencing capacity that has been developed. In FY 2002, the highest priority will be to finish the human genome, scheduled for completion in 2003. Beyond the human sequence, future large-scale sequencing will be aimed at developing data that will assist in interpreting the human sequence. One of the most efficient ways to do this is to obtain genomic sequence from related organisms. Because functionally important sequences are conserved, comparison between human and other organisms will help to identify many important features, including genes, gene structure, and regulatory and other non-coding functional elements.

Genomic Technology

The complexity of interpreting the genome requires that NHGRI redouble its efforts to develop new and more effective technologies for genomics research.

For example, comparative genomic sequencing is needed to identify genes, predict their functions, and understand how they are regulated. The genomic sequences of many organisms, at different evolutionary distances from humans, are needed for these tasks. To collect this vast amount of genomic sequence information from other organisms, even more cost-effective DNA sequencing technologies must be developed.

To map genes contributing to common diseases such as heart disease and diabetes, the technology for scoring large numbers of DNA sequence variants (single nucleotide polymorphisms, SNPs) in large numbers of individuals needs to become high throughput and 100 times cheaper. Until such technology becomes available, a good first step to find which SNPs may be involved in a disease is to compare SNP frequencies in a pooled sample of individuals affected by a disease with their frequencies in a pooled sample of unaffected individuals. SNPs that differ in frequency between the groups can then be followed up individually. The technology for determination of SNP frequencies in pooled DNA samples will receive special attention in FY 2002.

Genetic Variation Research and Health Disparities

The 1998 five year plan for the U.S. Human Genome Project (HGP) states that the Ethical, Legal and Social Implications (ELSI) Research Program should: “examine the issues surrounding the completion of the human DNA sequence and the study of human genetic variation” and “explore how socioeconomic factors, gender, and concepts of race and ethnicity influence the use, understanding, and interpretation of genetic information, the use of genetic services, and the development of policy.”

In FY 2002 this initiative on human genetic variation will be expanded to focus more specifically on how information generated by genetic variation research is likely to effect the ways in which culturally and socioeconomically diverse individuals and groups understand, access and use genetic information and health services, which in turn may affect health disparities. Topics to be addressed include: how will knowledge regarding genetics and its relationship to race and ethnicity be perceived and understood by people from different racial and ethnic groups?

- how will the identification of genetic differences or similarities between people from various racial and ethnic groups affect the health behaviors of members of those groups?
- how will health professionals' perceptions and beliefs about genetic variation likely influence their interactions with patients from diverse groups?
- what effect, if any, will continued developments in pharmacogenomic research aimed at identifying aggregate differences in drug response between members of different racial and ethnic groups have on health disparities?
- what effect will differential access to genetic predictive, diagnostic, and therapeutic interventions by members of culturally and socio-economically diverse groups likely have on health disparities?
- will increased public awareness and understanding of genetic variation generally help to *reduce* or to *exacerbate* existing health disparities?

Databases of genomic information

The large amount of genetic information being produced is most accessible and useful to researchers when it is in public databases. A major initiative at NHGRI is to expand and improve current databases and create new ones when the need arises.

In FY 2001, NHGRI initiated the development of a database for the roundworm (*C. elegans*) as well as the development of standard database modules that can be adapted for use by any organism community. These efforts will be expanded in FY 2002. In addition, NHGRI plans to evaluate and expand databases containing protein information. Another new initiative will be intensive annotation of the human sequence with biological information about the genes contained in it.

A major source of information on human gene function comes from studies on model organisms. NHGRI will expand its support of databases on model organisms, especially mouse, fly, worm, and yeast. NHGRI is supporting development of a novel annotation system called Gene Ontology (GO), which will allow ready comparison between organisms by employing a common classification scheme for genes, based on their function. These efforts will be expanded in FY 2002.

Training and Research Career Development Activities

Having the draft genome sequence available opens up the possibility of many types of genetic studies that were not possible prior to this time. It also creates a need to increase the number of scientists interested in studying diseases that are of interest to specific populations. Concomitant with this will be an increased need for researchers on the ethical, legal and social issues (ELSI) arising from such research.

It is anticipated that three new activities will evolve in FY 2002: (1) a training initiative in genomics research targeted to underrepresented minority undergraduate and graduate students and postdoctoral fellows; (2) a career development initiative in genomics research related to

health disparities targeted to postdoctoral fellows and junior faculty; and (3) an interdisciplinary ELSI/genetics training initiative.

Therapeutics in Genetics and Molecular Medicine

The intramural program is launching a new initiative in Therapeutics in Genetics and Molecular Medicine. There is a search underway to recruit investigators with the potential to establish programs that will lead to innovative approaches to therapy. Investigators will conduct basic and clinical research in areas such as biochemical genetics, proteomics, therapies for Mendelian or complex disorders, bioinformatics, and genome analysis.

Intramural Behavioral and Social Sciences Research

The NHGRI Intramural program will also be expanding research activities in behavioral and social sciences in human genetics. Efforts are underway to recruit a senior investigator to direct this initiative. The goal is to identify an outstanding leader to establish and direct a world-class behavior and social science research program within NHGRI. The investigator will have access to both the NIH Clinical Research Center for patient-oriented investigation and extensive interactions with NHGRI scientists conducting state-of-the-art genetics and genomics research.

INNOVATIONS IN MANAGEMENT AND ADMINISTRATION

NIH Collaborative Resource Centers

The NHGRI manages three NIH-wide scientific service centers that are supported by various NIH Institutes and Centers using a variety of funding mechanisms, such as reimbursable agreements, co-funding of grants, fee for services, etc. These scientific resource centers provide efficient and cost effective sequencing, genotyping, and genome-wide expression analysis for a variety of different intramural and extramural scientists supported by the NIH.

The *NIH Intramural Sequencing Center* provides NIH intramural investigators access to production-scale DNA sequencing and sequence analysis. Since its inception in 1997, the NIH Intramural Sequencing Center has performed 47 projects submitted by investigators from 13 NIH institutes. The *Center for Inherited Disease Research* provides high-throughput genotyping services, study design advice, sophisticated data warehousing technologies and database assistance to investigators attempting to identify gene variants involved in human disease. This Center is a joint effort supported by twelve NIH institutes with the NHGRI serving as the lead agency. A more detailed description of the scientific areas is provided earlier in this document.

The final resource center managed by the NHGRI is the *Microarray Technology Core.* This Core is a collaborative research effort between intramural scientists from many different NIH Institutes and programs and provides the specialized equipment and analysis for answering questions regarding gene expression and functional genomics.

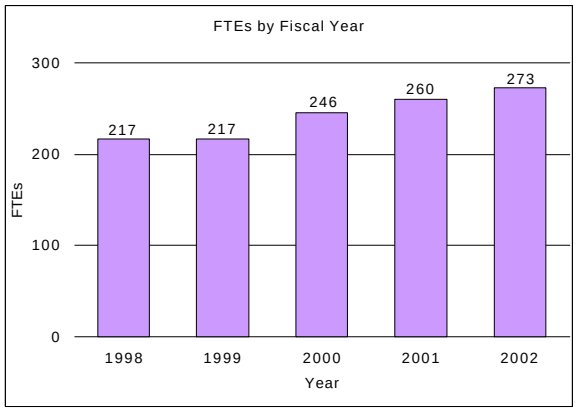
Data Safety and Monitoring Boards

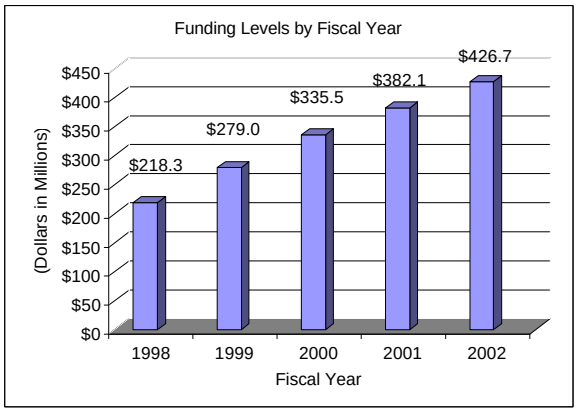
It is the policy of the NIH that each Institute and Center has a system for the appropriate oversight and monitoring of the conduct of clinical trials to ensure the safety of participants and the validity and integrity of the data for all NIH supported or conducted clinical trials. To ensure this oversight the NHGRI has established guidelines for use by the Institutional Review Board to determine the need for a data safety monitoring board for any protocol they review. This data safety monitoring board would be comprised of experts who monitor the clinical data and make findings based on the data. These experts help NHGRI investigators ensure patient safety and the highest quality of clinical studies.

BUDGET POLICY

The Fiscal Year 2002 budget request for the NHGRI is $426,739,000, including AIDS, an increase of $44,627,000 and 11.7 percent over the FY 2001 level, and $91,228,000 and 27.2 percent over FY 2000.

A five year history of FTEs and Funding Levels for NHGRI are shown in the graphs below:





One of NIH's highest priorities is the funding of medical research through research project grants (RPGs). Support for RPGs allows NIH to sustain the scientific momentum of investigator-initiated research while providing new research opportunities. The NIH Fiscal Year 2002 request provides average cost increases for competing RPGs equal to the Biomedical Research and Development Price Index (BRDPI), estimated at 4.3 percent. For NHGRI the average cost increase for competing RPGs is 2.7 percent, due to cycling of 2 large grants from competing to non-competing. Excluding these two grants, the average cost for competing research project grants will increase by 6.7 percent. Noncompeting RPGs will receive increases of 3 percent on average for recurring direct costs.

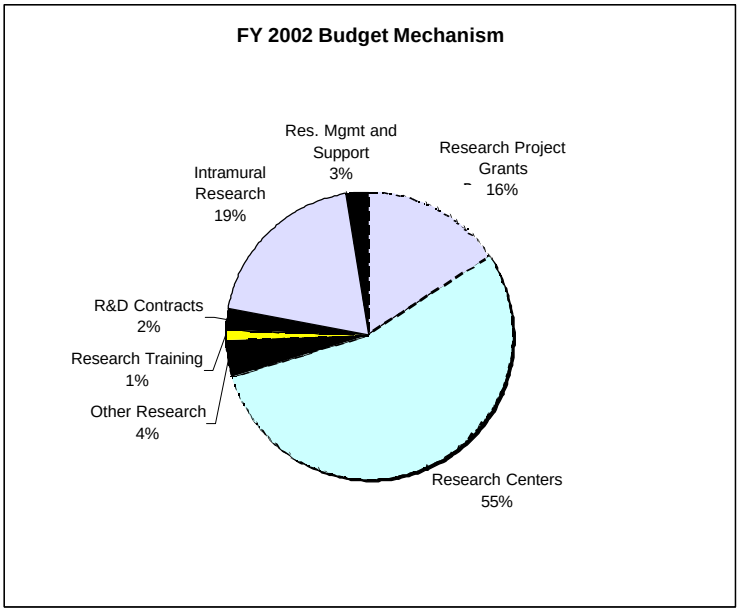
Promises for advancement in medical research are dependent on a continuing supply of new investigators with new ideas. In the Fiscal Year 2002 request, NHGRI will support 110 pre- and postdoctoral trainees in full-time training positions. An increase of 10 percent over Fiscal Year 2001 levels is provided for stipends and training-related expenses (e.g., health insurance, research supplies and equipment, and travel to scientific meetings.)

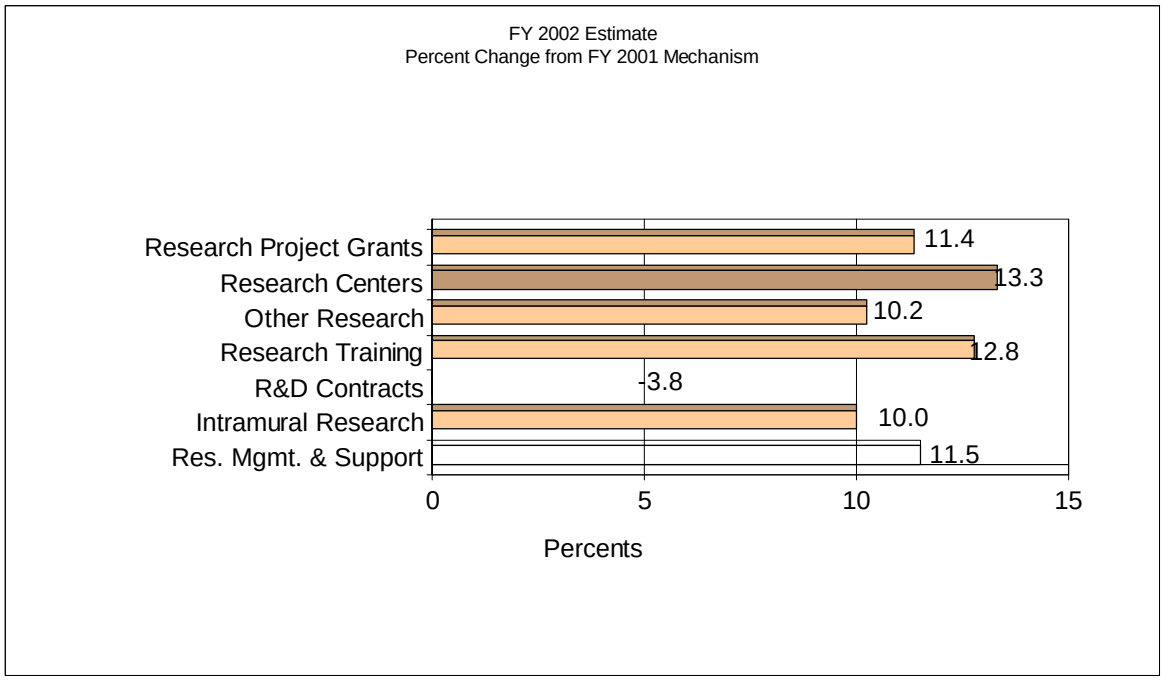
The Fiscal Year 2002 request includes funding for 28 research centers, 51 other research grants, including 21 research career awards, and 12 R&D contracts. The R&D contracts mechanism also includes support for 2 contracts for the Extramural Loan Repayment Programs.

The Research Mangement and Support budget activity will increase by 11.5 percent. This increase will provide for increases in FTE's and related expenses. NHGRI anticipates increases in staff to manage the growing portfolio of grants in the areas of computational biology, comparative sequencing, and technology development. NHGRI also expects to add additional

policy and outreach staff in FY 2002. These individuals are needed to help reach out more effectively to the various racial and ethnic groups that make up the U.S. population to understand their concerns and engage them in discussions.

The mechanism distribution by dollars and percent change are displayed below:





NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Budget Mechanism

MECHANISM	FY 2000 Actual		FY 2001 Estimate		FY 2002 Estimate	
Research Grants:	No.	Amount	No.	Amount	No.	Amount
Research Projects:						
Noncompeting	83	$46,824,000	78	$37,414,000	80	$46,503,000
Administrative supplements	(13)	6,442,000	(4)	1,000,000	(4)	1,200,000
Competing:						
Renew al	11	4,959,000	17	7,128,000	15	6,405,000
New	32	9,528,000	39	13,838,000	34	12,433,000
Supplements	0	0	0	0	0	0
Subtotal, competing	43	14,487,000	56	20,966,000	49	18,838,000
Subtotal, RPGs	126	67,753,000	134	59,380,000	129	66,541,000
SBIR/STTR	26	6,970,000	35	8,030,000	38	8,527,000
Subtotal, RPGs	152	74,723,000	169	67,410,000	167	75,068,000
Research Centers:						
Specialized/comprehensive	14	157,165,000	18	191,062,000	20	214,383,000
Clinical research	0	0	0	0	0	0
Biotechnology	6	9,464,000	6	9,552,000	8	12,945,000
Comparative medicine	0	0	0	0	0	0
Research Centers in Minority Institutions	0	75,000	0	0	0	0
Subtotal, Centers	20	166,704,000	24	200,614,000	28	227,328,000
Other Research:						
Research careers	13	1,672,000	19	3,200,000	21	3,605,000
Cancer education	0	0	0	0	0	0
Cooperative clinical research	0	0	0	0	0	0
Biomedical research support	0	0	0	0	0	0
Minority biomedical research support	0	0	0	0	0	0
Other	30	10,766,000	26	12,917,000	30	14,161,000
Subtotal, Other Research	43	12,438,000	45	16,117,000	51	17,766,000
Total Research Grants	215	253,865,000	238	284,141,000	246	320,162,000
Training:	FTTPs		FTTPs		FTTPs	
Individual awards	21	716,000	26	1,009,000	25	1,116,000
Institutional awards	54	2,077,000	81	3,208,000	85	3,639,000
Total, Training	75	2,793,000	107	4,217,000	110	4,755,000
Research & development contracts	16	6,604,000	11	10,531,000	12	10,133,000
(SBIR/STTR)	(0)	(0)	(0)	(0)	(0)	(0)
	FTEs		FTEs		FTEs	
Intramural research	182	64,197,000	192	73,721,000	203	81,094,000
Research management and support	64	8,052,000	68	9,502,000	70	10,595,000
Cancer prevention & control	0	0	0	0	0	0
Construction		0		0		0
Total, NHGRI	246	335,511,000	260	382,112,000	273	426,739,000
(Clinical Trials)		(6,814,000)		(7,473,000)		(8,185,000)

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Budget Authority by Activity
(dollars in thousands)

ACTIVITY	FY 2000 Actual		FY 2001 Estimate		FY 2002 Estimate		Change	
	FTEs	Amount	FTEs	Amount	FTEs	Amount	FTEs	Amount
Extramural Research:								
Human Geonome Research		$263,262		$298,889		$335,050		$36,161
Subtotal, Extramural research		263,262		298,889		335,050		36,161
Intramural research	182	64,197	192	73,721	203	81,094	11	7,373
Research management and support	64	8,052	68	9,502	70	10,595	2	1,093
Total	246	335,511	260	382,112	273	426,739	13	44,627

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Summary of Changes

2001 Estimated budget authority				$382,112,000
2002 Estimated budget authority				426,739,000
Net change				44,627,000
	2001 Current Estimate Base		Change from Base	
CHANGES	FTEs	Budget Authority	FTEs	Budget Authority
A. Built-in:				
1. Intramural research:				
a. Within grade increase		$20,026,000		$271,000
b. Annualization of January 2001 pay increase		20,026,000		185,000
c. January 2002 pay increase		20,026,000		553,000
d. Extra day of pay		20,026,000		79,000
e. Payment for centrally furnished services		13,146,000		1,315,000
f. Increased cost of laboratory supplies, materials, and other expenses		40,549,000		1,181,000
Subtotal				3,584,000
2. Research Management and Support:				
a. Within grade increase		5,945,000		103,000
b. Annualization of January 2001 pay increase		5,945,000		55,000
c. January 2002 pay increase		5,945,000		165,000
d. Extra day of pay		5,945,000		24,000
e. Payment for centrally furnished services		639,000		64,000
f. Increased cost of laboratory supplies, materials, and other expenses		2,918,000		0 132,000
Subtotal				543,000
Subtotal, Built-in				4,127,000

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Summary of Changes--continued

CHANGES	2001 Current Estimate Base		Change from Base	
	No.	Amount	No.	Amount
B. Program:				
1. Research project grants:				
a. Noncompeting	78	38,414,000	2	9,289,000
b. Competing	56	20,966,000	(7)	(2,128,000)
c. SBIR/STTR	35	8,030,000	3	497,000
Total	169	67,410,000	(2)	7,658,000
2. Centers	24	200,614,000	4	26,714,000
3. Other research	45	16,117,000	6	1,649,000
4. Research training	107	4,217,000	3	538,000
5. Research and development contracts	11	10,531,000	1	(398,000)
Subtotal, extramural				36,161,000
	FTEs		FTEs	
6. Intramural research	192	73,721,000	11	3,789,000
7. Research management and support	68	9,502,000	2	550,000
Subtotal, program		382,112,000		40,500,000
Total changes	260		13	44,627,000

NATIONL INSTITUTES OF HEALTH

National Human Genome Research Institute
Budget Authority by Object Class

		FY 2001 Estimate	FY 2002 Estimate	Increase or Decrease
Total compensable w orkyears:				
Full-time employment		260	273	13
Full-time equivalent of overtime and holiday hours		1	1	0
Average ES salary		$130,280	$135,035	$4,755
Average GM/GS grade		10.7	10.7	0.0
Average GM/GS salary		$55,315	$57,334	$2,019
Average salary, grades established by act of July 1, 1944 (42 U.S.C. 207)		$63,146	$65,451	$2,305
Average salary of ungraded positions		$218,357	$226,327	$7,970
	OBJECT CLASSES	FY 2001 Estimate	FY 2002 Estimate	Increase or Decrease
	Personnel Compensation:			
11.1	Full-Time Permanent	$10,492,000	$11,661,000	$1,169,000
11.3	Other than Full-Time Permanent	6,486,000	7,240,000	754,000
11.5	Other Personnel Compensation	754,000	831,000	77,000
11.8	Special Personnel Services Payments	3,503,000	3,694,000	191,000
11.9	**Total Personnel Compensation**	**21,235,000**	**23,426,000**	**2,191,000**
12.0	Personnel Benefits	4,735,000	5,284,000	549,000
13.0	Benefits for Former Personnel	1,000	1,000	0
	Subtotal, Pay Costs	**25,971,000**	**28,711,000**	**2,740,000**
21.0	Travel & Transportation of Persons	1,092,000	1,159,000	67,000
22.0	Transportation of Things	147,000	157,000	10,000
23.1	Rental Payments to GSA	2,000	2,000	0
23.2	Rental Payments to Others	185,000	197,000	12,000
23.3	Communications, Utilities & Miscellaneous Charges	698,000	735,000	37,000
24.0	Printing & Reproduction	75,000	84,000	9,000
25.1	Consulting Services	954,000	1,014,000	60,000
25.2	Other Services	7,293,000	8,241,000	948,000
25.3	Purchase of Goods & Services from Government Accounts	33,422,000	35,055,000	1,633,000
25.4	Operation & Maintenance of Facilities	426,000	437,000	11,000
25.5	Research & Development Contracts	2,500,000	2,757,000	257,000
25.6	Medical Care	568,000	590,000	22,000
25.7	Operation & Maintenance of Equipment	975,000	1,004,000	29,000
25.8	Subsistence & Support of Persons	0	0	0
25.0	**Subtotal, Other Contractual Services**	**46,138,000**	**49,098,000**	**2,960,000**
26.0	Supplies & Materials	10,874,000	12,062,000	1,188,000
31.0	Equipment	8,572,000	9,617,000	1,045,000
32.0	Land and Structures	0	0	0
33.0	Investments & Loans	0	0	0
41.0	Grants, Subsidies & Contributions	288,358,000	324,917,000	36,559,000
42.0	Insurance Claims & Indemnities	0	0	0
43.0	Interest & Dividends	0	0	0
44.0	Refunds	0	0	0
	Subtotal, Non-Pay Costs	**356,141,000**	**398,028,000**	**41,887,000**
	Total Budget Authority by Object	**382,112,000**	**426,739,000**	**44,627,000**

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Salaries and Expenses

OBJECT CLASSES	FY 2001 Estimate	FY 2002 Estimate	Increase or Decrease
Personnel Compensation:			
Full-Time Permanent (11.1)	$10,492,000	$11,661,000	$1,169,000
Other Than Full-Time Permanent (11.3)	6,486,000	7,240,000	754,000
Other Personnel Compensation (11.5)	754,000	831,000	77,000
Special Personnel Services Payments (11.8)	3,503,000	3,694,000	191,000
Total Personnel Compensation (11.9)	**21,235,000**	**23,426,000**	**2,191,000**
Civilian Personnel Benefits (12.0)	4,735,000	5,284,000	549,000
Benefits to Former Personnel (13.0)	1,000	1,000	0
Subtotal, Pay Costs	**25,971,000**	**28,711,000**	**2,740,000**
Travel (21.0)	1,092,000	1,159,000	67,000
Transportation of Things (22.0)	147,000	157,000	10,000
Rental Payments to Others (23.2)	185,000	197,000	12,000
Communications, Utilities and Miscellaneous Charges (23.3)	698,000	735,000	37,000
Printing and Reproduction (24.0)	75,000	84,000	9,000
Other Contractual Services:			
Advisory and Assistance Services (25.1)	954,000	1,014,000	60,000
Other Services (25.2)	7,293,000	8,241,000	948,000
Purchases from Govt. Accounts (25.3)	25,391,000	27,679,000	2,288,000
Operation & Maintenance of Facilities (25.4)	426,000	437,000	11,000
Operation & Maintenance of Equipment (25.7)	975,000	1,004,000	29,000
Subsistence & Support of Persons (25.8)	0	0	0
Subtotal Other Contractual Services	**35,039,000**	**38,375,000**	**3,336,000**
Supplies and Materials (26.0)	10,869,000	12,057,000	1,188,000
Subtotal, Non-Pay Costs	48,105,000	52,764,000	4,659,000
Total, Administrative Costs	**74,076,000**	**81,475,000**	**7,399,000**

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute
Significant Items in House and Senate Appropriations Committee Reports

FY 2001 House Appropriations Committee Report Language (H. Report 106-645)

Item

Diabetes - The Committee urges NHGRI, in collaboration with NIDDK and NICHD, to enhance research to identify the genes associated with juvenile, or Type I, diabetes. The Director of the Institute should be prepared to testify on this issue at the fiscal year 2002 appropriations hearing. (p. 88)

Action to be Taken

During fiscal year 2000 Human Genome Project scientists achieved an historic scientific milestone - completion of the initial reading of the text of our genetic instruction book. All this text, representing 90 percent of the 3.1 billion base pairs of the human genome, is freely available in public databases to all scientists with an Internet connection. Far from just a seminal scientific milestone, the achievement marked the unofficial starting point for a transformation in biomedical research that promises to revolutionize medicine. Determining the complete genetic blueprint of humans will greatly accelerate the identification of the genes underlying many human diseases, including common complex diseases such as diabetes. Identifying those genes is the first step to a more profound understanding of the biological basis of disease and this, in turn, will lead to much more effective and inexpensive ways to diagnose, treat, and prevent disease.

In fiscal year 2001, the NHGRI and its HGP partners will press ahead to complete the human sequence in order to close all gaps in the initial reading of the genome and ensure the information is as complete and accurate as technically possible. In addition, the HGP is developing additional tools for helping scientists analyze and understand the vast text of the human genome. Two activities are underway in FY 2001 that should enhance the success of work to identify genes involved in type I diabetes. These include the development of a catalog of common genetic variants (known as a SNPs catalog) and the completion of the initial reading of the mouse genome sequence. In collaboration with the NHLBI, an additional genome sequence effort is getting underway with the laboratory rat, for which animal models are also available that are highly relevant to diabetes. All these research tools will help investigators more quickly sift through the 3.1 billion bases of human DNA and isolate the genetic factors relating to disease, including the work of the NIDDK and NICHD to identify the genetic factors involved in type I Diabetes. The NHGRI will make all genomic tools developed by the HGP available to researchers for use in understanding human disease and improving their prevention, diagnosis and treatment.

Item

Renal Genomics - The Committee encourages NHGRI, in collaboration with NIDDK, to explore the feasibility of creating a kidney genome anatomy program. (p. 88)

Action to be Taken

At its core, the Human Genome Project is generating genomic information and developing tools for understanding the instructions encoded in human DNA. Having the sequence, or spelling, of the human genome and the tools to interpret its meaning, will enable biomedical researchers in all areas to answer questions about disease processes and help develop new strategies for their prevention, diagnosis and treatment. Last June, Human Genome Project participants announced that they had completed a "working draft" of the human genome and that the draft was freely available in public databases. While this marked a significant milestone of the Human Genome Project, much work remains to be done in order for disease investigators to make sense of the information. The working draft, while not as refined as the final version, includes sequence covering most of the genome and represents the raw data needed to find most of the human genes.

In FY 2001, the NHGRI will continue to support HGP scientists generating the information and tools necessary for understanding the human genome. The first priority is to "finish" the human sequence. The final human genome sequence will be highly refined, that is 99.99% accurate and is expected to be done by 2003. HGP partners will continue their work to complete sequencing the mouse and rat genomes, to develop a catalog of common variants of the human genome, and to improve computational and informatics tools.

Another important genomic resource supported by the NHGRI is the development of the Mammalian Gene Collection, or MGC. When a gene is active in a cell, the functional, or coding part of the gene is "transcribed" into a mRNA molecule as a precursor to making the protein. A particularly powerful material for studying gene expression and gene function, therefore, is cDNA, which represents the full-length, expressed mRNA transcript. The MGC's scientific goals are to produce publicly accessible cDNA clone collections that contain full-length copies of all genes from all tissues, including the kidney, to sequence these cDNAs, to develop the associated informatics tools, and to create a publicly accessible website to provide up-to-date information to the research community. This project was initiated in FY 1999 by the NHGRI and the National Cancer Institute (NCI), and involves participation of 15 other NIH Institutes, including the NIDDK. Complete catalogs of genes from all tissues will be essential for thorough genetic and physiological analysis. The MGC program is designed to generate these critical resources, which will be widely used and of inestimable value to researchers studying the role of genes, and gene pathways, in disease, including diseases involving the kidney.

In addition, NHGRI intramural investigators will continue their work on the development and use of microarray technology in the study of human disease. Much of this work involves productive collaborations with numerous intramural scientists in multiple institutes of the NIH. The flood of data emerging about DNA and genes has required new ways of sorting through the

information to find the telling details that will illuminate how the cells that make up living organisms function. Microarrays are being used for many different applications. Some microarrays gauge how active different genes are in different kinds of cells; others let researchers track the molecular changes in tumor cells as cancer progresses. Capturing holistic views of changes within cells has begun to elucidate the signaling pathways that are altered and distinguish a cancerous cell from a non-cancerous cell. Such insights may provide the identification of early, presymptomatic changes in cells and thus, rational therapeutic targets for the treatment of cancer, diabetes, and other diseases.

All of these efforts by NHGRI supported HGP partners and intramural scientists will improve the ability of investigators to gather, organize, and analyze the vast amount of genomic data relevant for understanding the genetic factors involved in disease, including those involved in kidney disease. The NHGRI welcomes opportunities to assist researchers supported by other NIH institutes and centers in utilizing these powerful genomic tools in their study of human disease.

FY 2001 Senate Appropriations Committee Report Language (S. Rept. 106-293)

Item

Clearinghouse for rare and genetic disorders - The Committee is pleased to learn that the NHGRI and the ORD are working on establishing such a center and encourages them to move as expeditiously as possible to provide this important information service to the public and health professionals. (p. 168)

Action to be Taken

The NHGRI, in collaboration with the Office of Rare Diseases (ORD), is proceeding as expeditiously as possible on supporting an information clearinghouse for rare and genetic disorders. Based upon current progress, the NHGRI and the ORD expect to make an award to establish a Genetic and Rare Disease Information Center in FY 2001.

Item

Diabetes - The Committee urges the NHGRI to assist the NIDDK and NICHD in the consideration of a collaborative project to identify the genes associated with juvenile, or Type 1, diabetes. (p. 168)

Action to be Taken

During fiscal year 2000 Human Genome Project scientists achieved an historic scientific milestone - completion of the initial reading of the text of our genetic instruction book. All this text, representing 90 percent of the 3.1 billion base pairs of the human genome, is freely available in public databases to all scientists with an Internet connection. Far from just a seminal scientific milestone, the achievement marked the unofficial starting point for a transformation in

biomedical research that promises to revolutionize medicine. Determining the complete genetic blueprint of humans will greatly accelerate the identification of the genes underlying many human diseases, including common complex diseases such as diabetes. Identifying those genes is the first step to a more profound understanding of the biological basis of disease and this, in turn, will lead to much more effective and inexpensive ways to diagnose, treat, and prevent disease.

In fiscal year 2001, the NHGRI and its HGP partners will press ahead to complete the human sequence in order to close all gaps in the initial reading of the genome and ensure the information is as complete and accurate as technically possible. In addition, the HGP is developing additional tools for helping scientists analyze and understand the vast text of the human genome. Two activities are underway in FY 2001 that should enhance the success of work to identify genes involved in type I diabetes. These include the development of a catalog of common genetic variants (known as a SNPs catalog) and the completion of the initial reading of the mouse genome sequence. In collaboration with the NHLBI, an additional genome sequence effort is getting underway with the laboratory rat, for which animal models are also available that are highly relevant to diabetes. All these research tools will help investigators more quickly sift through the 3.1 billion bases of human DNA and isolate the genetic factors relating to disease, including the work of the NIDDK and NICHD to identify the genetic factors involved in type I Diabetes. The NHGRI will make all genomic tools developed by the HGP available to researchers for use in understanding human disease and improving their prevention, diagnosis and treatment.

Item

Kidney genome anatomy - The Committee encourages the NHGRI to assist the NIDDK in an effort to create a kidney gnome anatomy program. Given the NHGRI's expertise in areas such as gene sequencing, expression, and microarray technology, the Committee believes that NHGRI could provide valuable assistance to the NIDDK.

Action to be Taken

Please refer to page 29 of this document for the NHGRI response to this significant item regarding the kidney genome anatomy program.

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Authorizing Legislation

	PHS Act/ Other Citation	U.S. Code Citation	2000 Amount Authorized	2001 Estimate	2002 Amount Authorized	2002 Budget Estimate
Research and Investigation	Section 301	42§241	Indefinite	$377,895,000	Indefinite	$421,984,000
National Human Genome Research Institute	Section 485B	42§287	Indefinite		Indefinite	
National Research Service Awards	Section 487(d)	42§288	a/	4,217,000	b/	4,755,000
Total, Budget Authority				382,112,000		426,739,000

a/ Funding provided under the Departments of Labor, Health and Human Services, Education, and Related Agencies Appropriations Act, 2001 (P.L. 106-554).

b/ Reauthorizing legislation will be submitted.

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Appropriation History

Fiscal Year	Budget Estimate to Congress	House Allowance	Senate Allowance	Appropriation 1/
1993	$110,429,000	$107,217,000	107,217,000	$106,239,000 2/
1994	134,549,000	119,030,000	131,925,000	127,112,000
1995 3/	152,010,000	151,878,000	151,878,000	151,518,000 4/
Rescission				(331,000)
1996	166,678,000 3/	170,041,000	163,943,000 3/	169,041,000
Rescission				(266,000)
1997	177,788,000 3/	189,267,000	180,807,000 3/	189,657,000 5/
1998	202,197,000 3/	211,772,000	218,851,000	217,704,000
1999	236,275,000 3/6/	246,111,000	249,891,000	264,892,000
Rescission				(185,000)
2000	271,536,000 3/	308,012,000	337,322,000	337,322,000
Rescission				(1,795,000)
2001	353,427,000 3/	386,410,000	385,888,000	382,384,000
Rescission				(192,000)
2002	426,739,000			

1/ Reflects enacted supplementals, rescissions and reappropriations.
2/ Excludes enacted administrative reductions of $978,000.
3/ Excludes funds for HIV/AIDS research activities consolidated in the NIH Office of AIDS research.
4/ Excludes enacted administrative reductions of $161,000.
5/ Excludes enacted administrative reductions of $128,000.
6/ Excludes reductions of $721,000 for the budget amendment for bioterrorism.

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

Detail of Full-Time Equivalent Employment (FTEs)

OFFICE/DIVISION	FY 2000 Actual	FY 2001 Estimate	FY 2002 Estimate
Office of the Director	7	7	7
Office of Administrative Management	22	23	23
Office of Policy and Public Affairs	10	14	15
Division of Intramural Research	182	192	203
Division of Extramural Research	25	24	25
Total, NHGRI	246	260	273
FTEs supported by funds from Cooperative Research and Development Agreements	(1)	(3)	(1)

FISCAL YEAR	Average GM/GS Grade
1998	10.9
1999	10.9
2000	10.8
2001	10.7
2002	10.7

NATIONAL INSTITUTES OF HEALTH
National Human Genome Research Institute
Program Administration

Detail of Positions

GRADE	FY 2000 Actual	FY 2001 Estimate	FY 2002 Estimate
ES-6			
ES-5	2	2	2
ES-4			
ES-3	1	1	1
ES-2			1
ES-1			
Subtotal	3	3	4
Total - ES Salary	$387,225	$390,840	$441,070
GM/GS-15	23	23	23
GM/GS-14	13	13	14
GM/GS-13	23	24	25
GS-12	37	38	38
GS-11	20	22	23
GS-10	2	2	2
GS-9	28	28	29
GS-8	16	16	17
GS-7	11	12	12
GS-6	5	5	5
GS-5	4	4	4
GS-4	2	2	2
GS-3	3	3	3
GS-2	2	3	4
GS-1	0	0	0
Subtotal	189	195	201
Grades established by Act of July 1, 1944 (42 U.S.C. 207):			
Assistant Surgeon General			
Director Grade	1	1	1
Senior Grade	1	2	2
Full Grade			
Senior Assistant Grade			
Assistant Grade			
Co-Step			
Subtotal	2	3	3
Ungraded	85	93	97
Total permanent positions	158	164	171
Total positions, end of year	279	294	305
Total full-time equivalent (FTE) employment,end of year	246	260	273
Average ES level	ES-4	ES-4	ES-4
Average ES salary	$129,075	$130,280	$135,035
Average GM/GS grade	10.8	10.7	10.7
Average GM/GS salary	$55,218	$55,315	$57,334

NATIONAL INSTITUTES OF HEALTH

National Human Genome Research Institute

New Positions Requested

	FY 2002		
	Grade	Number	Annual Salary
Biological Lab Technician	GS-9	2	$38,001
Biologist	GS-11	2	45,977
Computer Specialist	GS-12	1	54,103
Program Analyst (Health Disparities)	GS-13	1	65,525
Computational Biologist	GS-14	1	77,430
Total Requested		7	